



Brevet d'invention

Certificat d'utilité

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 10 DEC. 2009

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

A handwritten signature in black ink, appearing to read 'M. Planche', enclosed within a large, stylized oval loop.

Martine PLANCHE

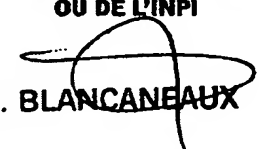
REQUÊTE EN DÉLIVRANCE 1/2

Remplir impérativement la 2ème page.

Cet imprimé est à remplir lisiblement à l'encre noire

08 540 W / 190600

<p>REMISE DES PIÈCES DATE 14 JUIN 2001 LIEU 75 INPI PARIS N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI 0107808 DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI 14 JUIN 2001</p>		<p>1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE</p> <p>BREESE-MAJEROWICZ 3 avenue de l'Opéra 75001 PARIS</p>	
<p>Vos références pour ce dossier (facultatif) 13052FR</p>			
<p>Confirmation d'un dépôt par télécopie <input type="checkbox"/> N° attribué par l'INPI à la télécopie</p>			
<p>2 NATURE DE LA DEMANDE</p>		<p>Cochez l'une des 4 cases suivantes</p>	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
<p><i>Demande de brevet initiale</i> N° _____ Date ____/____/____</p> <p><i>ou demande de certificat d'utilité initiale</i> N° _____ Date ____/____/____</p>			
<p>Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i></p>		<p><input type="checkbox"/> N° _____ Date ____/____/____</p>	
<p>3 TITRE DE L'INVENTION (200 caractères ou espaces maximum)</p> <p>METHODE D'IDENTIFICATION DE MOTIFS ET/OU DES COMBINAISONS DE MOTIFS PRESENTANT UN ETAT BOOLEEN DE MUTATION PREDETERMINEE DANS UN ENSEMBLE DE SEQUENCES ET SES APPLICATIONS</p>			
<p>4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE</p>		<p>Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»</p>	
<p>5 DEMANDEUR</p>		<p><input checked="" type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»</p>	
Nom ou dénomination sociale		CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS -	
Prénoms			
Forme juridique			
N° SIREN		
Code APE-NAF		
Adresse	Rue	3 rue Michel-Ange	
	Code postal et ville	75794	PARIS Cedex 16
Pays		France	
Nationalité		Française	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			

REMISE DES PIÈCES DATE 14 JUIN 2001 LIEU 75 INPI PARIS N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI 0107808		Réservé à l'INPI		DB 540 W / 190600	
Vos références pour ce dossier : <i>(facultatif)</i>			13052FR		
6 MANDATAIRE					
Nom			BREESE		
Prénom			Pierre		
Cabinet ou Société			BREESE-MAJEROWICZ		
N °de pouvoir permanent et/ou de lien contractuel					
Adresse	Rue	3 avenue de l'Opéra			
	Code postal et ville	75001	Paris		
N° de téléphone <i>(facultatif)</i>		01 47 03 67 77			
N° de télécopie <i>(facultatif)</i>		01 47 03 67 78			
Adresse électronique <i>(facultatif)</i>		office@breese.fr			
7 INVENTEUR (S)					
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée			
8 RAPPORT DE RECHERCHE		Uniquement pour une demande de brevet (y compris division et transformation)			
Établissement immédiat ou établissement différé		<input checked="" type="checkbox"/> <input type="checkbox"/>			
Paiement échelonné de la redevance		Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non			
9 RÉDUCTION DU TAUX DES REDEVANCES		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention <i>(joindre un avis de non-imposition)</i> <input type="checkbox"/> Requête antérieurement à ce dépôt <i>(joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence) :</i>			
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes		1			
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire) BREESE Pierre 921038			VISA DE LA PRÉFECTURE OU DE L'INPI  M. BLANCANEUX		

REQUÊTE EN DÉLIVRANCE

Page suite N° 1.../1...

Remise des pièces
DATE **14 JUIN 2001**
LIEU **75 INPI PARIS**

N° D'ENREGISTREMENT **0107808**
NATIONAL ATTRIBUÉ PAR L'INPI

Cet imprimé est à remplir lisiblement à l'encre noire

DB 829 W / 260899

Vos références pour ce dossier (facultatif)		13052FR	
4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation	N°
		Date	
		Pays ou organisation	N°
		Date	
		Pays ou organisation	N°
		Date	
5 DEMANDEUR			
Nom ou dénomination sociale		INSTITUT PASTEUR	
Prénoms			
Forme juridique			
N° SIREN			
Code APE-NAF			
Adresse	Rue	25-28 rue du Docteur Roux	
	Code postal et ville	75724	PARIS Cedex 15
Pays		France	
Nationalité		France	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			
5 DEMANDEUR			
Nom ou dénomination sociale			
Prénoms			
Forme juridique			
N° SIREN			
Code APE-NAF			
Adresse	Rue		
	Code postal et ville		
Pays			
Nationalité			
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire)		VISA DE LA PRÉFECTURE OU DE L'INPI	
BREESE Pierre 921038		M. BLANCANEUX	

METHODE D'IDENTIFICATION DE MOTIFS ET/OU DES COMBINAISONS
DE MOTIFS PRESENTANT UN ETAT BOOLEEN DE MUTATION
PREDETERMINEE DANS UN ENSEMBLE DE SEQUENCES ET SES
APPLICATIONS.

5

L'invention appartient au domaine d'analyse des
séquences de nucléotides et/ou d'acides aminés composant
les organismes vivants, en particulier l'analyse de
mutations particulières que lesdites séquences peuvent
10 présenter.

Elle concerne des méthodes d'identification et
de sélection de fragments de séquences d'acides nucléiques
ou de protéines constitués par et/ou comprenant des motifs
15 présentant des caractéristiques de mutabilité spécifiques,
elle concerne également des compositions pharmaceutiques
contenant lesdites fragments pour la préparation de
médicaments utiles pour le traitement et/ou la prévention,
de pathologies humaines, animales et/ou végétales ou pour
20 la préparation de cibles thérapeutiques utiles pour le
criblage de composés thérapeutiques.

On connaît que des mutations induites dans les
séquences sauvages d'organismes pathogènes sont par
25 exemple, responsables des mécanismes d'échappement
thérapeutique, c'est à dire de la capacité des organismes
pathogènes, viraux ou bactériens, à résister à un
traitement thérapeutique. Les séquences nucléotidiques
et/ou polypeptidiques des souches mutantes desdits
30 organismes présentent en effet des mutations particulières
par rapport aux séquences nucléotidiques ou polypeptidiques
des souches sauvages

De telles mutations sont également déterminantes de changements fonctionnels des gènes ou des protéines qui ont pour conséquence l'altération de nombreux processus biologiques, tels que le déclenchement de la réponse immune, l'infectivité des virus, l'apparition de cancers, etc.

On connaît, par exemple, que l'information génétique du virus de l'immunodéficience humaine (VIH), appartenant à la famille des rétrovirus, est supportée par deux molécules d'ARN. Lors de l'infection, l'intégration du génome viral à celui de la cellule hôte ne peut donc se faire directement. La synthèse préalable d'une copie d'ADN à partir de l'ARN génomique du virus est une étape déterminante du cycle infectieux. L'enzyme responsable de cette transcription inverse est une protéine appelée Reverse Transcriptase (RT). La faible fidélité reverse-transcriptionnelle de cette dernière confère au virus une grande variabilité génomique. On estime que chez un individu séropositif non traité, une mutation apparaît par réplication, et donc pour les dix milliards de virus produits par jour, il y a 10 milliards de mutation nouvelles. Cette mutation peut entraîner une résistance à un ou plusieurs antirétroviraux et ainsi générer des "souches" plus virulentes car de plus en plus résistantes.

Face à cette problématique, les praticiens prescrivent des traitements très lourds, tels que la trithérapie à long terme, depuis peu la quadrithérapie et peut être plus à l'avenir, profitant de l'absence de virus résistant qui caractérise en général les patients non encore traités et infectés par une seule forme du virus. Ces traitements provoquent alors une forte diminution de la

charge virale, considérée comme la quantité de particules virales circulant dans le sang, le nombre de mutants viraux qui est directement proportionnel à la charge virale, diminue également, réduisant ainsi les risques d'échappement thérapeutique.

Malheureusement, ces traitements extrêmement lourds s'accompagnent de nombreux effets secondaires. Ils nécessitent, en outre, une compliance parfaite qui, lorsqu'elle n'est pas respectée, s'accompagne presque systématiquement de l'émergence de souches résistantes. Ces résistances sélectionnées sous la pression des antirétroviraux sont à l'origine de la plupart des échappements thérapeutiques.

Ainsi, alors que le choix d'une combinaison d'antirétroviraux apparaît comme fondamental, l'association optimisée de ces derniers ne semble pas évidente. Outre les problèmes multiples posés par les résistances que nous venons de décrire, l'incompatibilité de certaines associations médicamenteuses et le nombre toujours croissant de molécules antirétrovirales rendent le travail des praticiens de plus en plus ardu.

A l'heure actuelle les médecins disposent d'une vingtaine de composés thérapeutiques, essentiellement dirigés contre deux protéines virales, la reverse transcriptase et la protéase. Les traitements les plus usuels sont les trithérapies. On en dénombre 252 possibles lorsqu'on ne considère que les associations les plus courantes. Ces calculs sont statistiques et ne prennent pas en compte les différentes incompatibilités médicamenteuses. De plus, l'apparition de nouveaux principes actifs issus de

la recherche pharmaceutique aura pour conséquence directe de compliquer encore le problème du choix de la combinaison médicamenteuse.

5 L'activité d'autres organismes pathogènes est tout aussi préoccupante, le virus de la grippe a été responsable de 20 millions de décès durant le XX^{ème} siècle et le virus Ebola émerge de façon alarmante. Les hépatites A, B, C, D et E constituent de véritables priorités de
10 santé publique, de par leur état booléen et leur gravité potentielle.

Or, dans tous les cas il y a un vide thérapeutique et vaccinal qui s'accroît chaque année à
15 cause de la grande mutabilité des génomes viraux: spécialement des rétrovirus, virus à ARN tels que VIH, grippe, Ebola, hépatite C, etc.

Plusieurs approches ont été proposées pour
20 tenter de résoudre ces problèmes de multirésistance liée à la haute mutabilité de certains organismes pathogènes, ainsi, par exemple, la société Virco Tibotech, a développé une méthode gérée par un logiciel qui permet la comparaison d'un génotype donné à toute une banque de séquences VIH. Il
25 définit ensuite la liste des résistances possibles aux composés antirétroviraux.

Aussi, certains sites web, tel que celui de la Los-Alamos Library (<http://hiv-web.lanl.gov/>) fournissent
30 un grand nombre de données concernant les alignements de séquences protéiques de VIH ainsi que les mutations s'y rapportant.

De même, plusieurs publications de Ribeiro et al., divulguent des méthodes mettant en oeuvre de calculs d'état booléens d'apparition de mutants résistants en utilisant des calculs mathématiques assez complexes.

5

Aussi, des méthodes visant à identifier des mutations des motifs constituants des séquences nucléotidiques ou polypeptidiques ont été développées, par exemple, celles qui ont permis, dans les années 80 de classer les gènes des immunoglobulines en classes et sous-classes, comportant des domaines constants et des domaines variables en fonction de la variabilité de motifs des différentes séquences qui les composent.

10

15

20

Cependant ces méthodes ne permettent pas d'identifier des motifs dont la possibilité de mutation est prédéterminée par rapport à l'ensemble de séquences analysée. Dans le cadre de la présente invention, cette possibilité de mutation correspond à un état booléen de ladite mutation.

25

30

La méthode de l'invention a pour objet l'identification de plusieurs motifs dont l'état booléen de mutation relative, par rapport à un ensemble de séquences données, est prédéterminée. Cette méthode est basée sur l'identification soit des motifs ou de combinaisons de motifs n'ayant jamais muté simultanément, soit de motifs ou de combinaison de motifs ayant muté simultanément, au moins une fois sur au moins une des séquences de l'ensemble et n'ayant pas muté sur les autres séquences dudit ensemble.

La présente invention constitue un nouvel outil pour permettre de trouver des solutions plus durables lors

des traitements thérapeutiques des pathologies impliquant des organismes pathogènes ou des gènes humains, présentant un haut degré de mutabilité.

5 L'invention a aussi pour objet l'utilisation des séquences constituées ou comprenant lesdits motifs et/ou des combinaisons des motifs ainsi identifiés pour la préparation de médicaments et/ou de vaccins utiles pour le traitement ou la prévention de pathologies humaines,
10 animales ou végétales, la préparation de cibles thérapeutiques utiles pour le criblage de tels médicaments, l'arrimage (docking) d'un médicament sur sa cible, la conception de nouvelles méthodes d'aide au diagnostic, où par exemple, le choix d'un ou plusieurs composés
15 thérapeutiques s'effectuerait en fonction de la mutabilité des organismes pathogènes à l'origine de la maladie d'un patient donné.

Au sens de la présente invention on entend par
20 motif un nucléotide susceptible de faire partie d'une séquence d'acide nucléique ou d'un oligonucléotide synthétique, désigné ci-après par son code unicaractère: A, G, C, T ou U, correspondant à la nomenclature de la base respective (adénine A, guanine G, cytosine C, ou thymine T
25 dans l'ADN, ou uracile U dans l'ARN) dont ils sont constitués.

On entend également par motif un acide aminé, quelle que soit sa configuration, susceptible de faire partie d'une protéine ou d'un peptide naturel ou
30 synthétique, désigné par son code unicaractère tels que par exemple, ceux représentés dans le tableau ci-dessus.

Code des acides aminés

Code	aa
A	Alanine
C	Cystéine
D	Acide Aspartique
E	Acide Glutamique
F	Phénylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Méthionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Thréonine
V	Valine
W	Tryptophane
Y	Tyrosine

On entend par séquence, tout enchaînement de motifs tels que ci-dessus définis, susceptible de constituer une séquence d'un acide nucléique ou un fragment de celui-ci d'un organisme vivant ou une séquence d'une protéine ou un fragment de celle-ci d'un organisme vivant y compris les séquences sauvages, les séquences mutantes ou encore, des séquences artificielles analogues de celles-ci obtenus par synthèse chimique ou biologique selon des méthodes connues de l'homme du métier. A titre d'exemple, et de manière non limitative, on entend par séquence contenant de tels motifs, un groupe de gènes, un gène, ou un fragment de celui-ci, un groupe de protéines, une protéine ou un fragment de celle-ci.

On entend par variante d'une séquence toute séquence différant de la séquence originale ou sauvage par au moins un motif.

Ainsi l'invention a pour objet l'identification de motifs n'ayant jamais muté simultanément parmi tous les membres d'un ensemble de séquences. L'identification de tels motifs est un enjeu majeur des nouveaux développements pharmacologiques, tant au niveau des cibles thérapeutique comme au niveau de composés thérapeutiques recherchés, notamment dans le cadre de résistances et de multirésistances développées par des organismes pathogènes nocifs tant pour l'espèce animal comme pour l'espèce végétal.

L'invention concerne aussi l'utilisation de ces fragments de séquences constitués par et/ou comprenant des motifs n'ayant jamais muté simultanément pour la préparation de cibles thérapeutiques utiles pour le criblage de médicaments ainsi que pour la préparation de vaccins dirigés contre des organismes pathogènes et en particulier contre des organismes pathogènes présentant un degré élevé de mutabilité.

L'invention concerne enfin l'utilisation de séquences constituées par et/ou comprenant des motifs n'ayant jamais muté simultanément pour la préparation des composés utiles pour la prévention et le traitement de pathologies humaines et/ou animales et en particulier des pathologies dont des gènes responsables présentent un haut degré de mutabilité.

L'utilisation de fragments de séquences particulières desdits organismes pathogènes, constitués par et/ou comprenant lesdits motifs qui n'ont jamais muté

simultanément en tant que composés thérapeutiques permettra, entre autres, de:

- Diminuer l'apparition de résistances aux traitements thérapeutiques;

5 - Stabiliser la santé du patient sur le long terme en permettant l'utilisation des médicaments disponibles sur le marché plus longtemps;

- Eviter l'apparition de maladies opportunistes ce qui diminuera le coût global du traitement;

10 - Diminuer la durée et le coût des investissements en recherche et développement dans l'industrie pharmaceutique.

15 La présente invention propose donc un nouvel outil pour optimiser le choix des traitements thérapeutiques dirigés contre des organismes pathogènes à fort taux de mutabilité ou contre des pathologies dues à l'apparition de mutations..

20 La méthode d'identification de motifs de l'invention consiste à comparer un sous-ensemble de variantes d'une même séquence nucléotidique ou polypeptidique d'un organisme pathogène donné, au moyen d'une séquence de référence, par exemple une séquence
25 consensus, et à identifier lors de cette comparaison, les motifs desdites séquences qui ne mutent jamais simultanément ou les motifs qui mutent simultanément au moins une fois sur au moins une des séquences du sous-ensemble et ne mutent pas sur les autres séquences dudit
30 sous-ensemble.

Plus précisément l'invention a pour objet une méthode d'identification d'un motif ou d'une combinaison de

motifs présentant un état booléen de mutation prédéterminée dans un ensemble de séquences, caractérisée en ce qu'elle comprend les étapes suivantes;

a) l'alignement de séquences de motifs ordonnés représentées par leur code unicaractère,

b) la comparaison d'une séquence de référence à l'ensemble de séquences alignées à l'étape (a),

c) l'identification des motifs n'ayant jamais muté simultanément ou des motifs ayant muté simultanément au moins une fois sur au moins une des séquences de l'ensemble et n'ayant pas muté sur les autres séquences dudit ensemble.

Selon un premier mode de réalisation de l'invention, le motif ou la combinaison de motifs à identifier est un nucléotide ou une combinaison de nucléotides et le sous-ensemble de séquences peut être extrait d'une banque de données d'acides nucléiques.

Selon un deuxième mode de réalisation, le motif ou la combinaison de motifs à identifier est un acide aminé ou une combinaison d'acides aminés et le sous-ensemble de séquences peut être extrait d'une banque de données de polypeptides et/ou de protéines.

L'alignement des séquences peut-être effectué selon toute méthode d'alignement connue de l'homme du métier.

Par exemple, lorsque le nombre de séquences du sous-ensemble que l'on utilise est inférieur à 100, on peut utiliser la méthode d'alignement Clustal W. (Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific

gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680).

Si le nombre de séquences à analyser est plus important, par exemple, supérieur à 100, l'alignement proposé par Clustal W est trop long et on peut alors avoir recours à un alignement itératif basé sur un modèle de Markov caché, ci après désigné HMM. (Sean Eddy. " Hidden Markov Models ", Curr.Opin.Struct.Biol. Vol.6, pages 361-365, 1966).

Dans ce dernier cas, il est crée, par exemple, un premier sous-ensemble de 100 séquences extraites de l'ensemble de séquences à analyser, auquel on applique la méthode de Clustal pour obtenir un premier alignement.

A partir de ce premier alignement, on crée un modèle de Markov caché (HMM) le modèle est éventuellement calibré afin de le rendre plus sensible, puis on ajoute audit premier alignement des nouvelles séquences qui seront à leur tour alignées en utilisant à nouveau HMM

Avantageusement, la séquence de référence de l'étape (b) est constituée par une séquence sauvage, ou par une séquence consensus comportant en position i le motif présent en position i dans un nombre prédéterminé des séquences de l'étape (a), par exemple dans plus de 30% desdites séquences et plus préférentiellement dans plus de 75% desdites séquences, ces valeurs pouvant être réglables selon les cas.

Avantageusement, l'étape (b) de comparaison de séquences de la méthode d'identification de l'invention consiste à:

- constituer une première matrice numérique A de dimensions $N \times M$ où N désigne le nombre de séquences et M désigne le nombre de motifs d'une des séquences dudit alignement, la valeur $A_{i,j}$ étant égale à une première valeur A1 [par exemple "0"] lorsque le motif de position i de la séquence j est muté par rapport au motif de position i de la séquence de référence, et égale à une deuxième valeur A2 [par exemple "1"] dans les autres cas,

- constituer deux matrices d'analyse B, C des mutations où ces matrices sont :

- une matrice B de couples non mutés, c'est-à-dire de couples qui ne mutent jamais simultanément, de dimension $M \times M$, la valeur $B_{i,k} = B_{k,i}$ étant égale :

- à une première valeur B1 [par exemple "0"] lorsque $A_{i,j} = A_{k,j} = A1$ quel que soit j allant de 0 à N,
- à une deuxième valeur B2 [par exemple "1"] dans les autres cas;

- une matrice C de couples mutés [c'est-à-dire de couples qui mutent soit toujours, soit jamais simultanément] de dimension $M \times M$, la valeur $C_{k,i} = C_{i,k}$ étant égale :

- à une deuxième valeur C1 [par exemple "1"] lorsque $A_{i,j} = A_{k,j}$ quel que soit j allant de 0 à N,
- à une première valeur C2 [par exemple "0"] dans les autres cas;

- à déterminer, pour un ensemble E de positions, un coefficient R_E dont la valeur est R_1 [par exemple "1"] lorsque toutes les valeurs $B_{i,k}$ sont égales à la deuxième valeur B2, quels que soient i et k appartenant à l'ensemble E desdites positions, où $i \leq k$.

- à déterminer, pour un ensemble F de positions, un coefficient R_F dont la valeur est R_1 [par

exemple "1"] lorsque toutes les valeurs $C_{i,k}$ sont égales à la deuxième valeur C_1 , quels que soient i et k appartenant à l'ensemble F desdites positions, où $i \neq k$.

5 Avantageusement, la matrice de couples mutés de l'invention permet d'identifier deux motifs ayant muté simultanément au moins une fois sur au moins une des séquences de l'ensemble et n'ayant pas muté sur les autres séquences dudit ensemble.

10 L'invention concerne également l'algorithme développé pour effectuer la comparaison des séquences contenant lesdits motifs et l'identification des motifs de celles-ci, soit ayant muté simultanément au moins une fois sur au moins une des séquences de l'ensemble et n'ayant pas muté sur les autres séquences dudit ensemble et consistant à :

20 - constituer une première matrice numérique A de dimensions $N \times M$ où N désigne le nombre de séquences et M désigne le nombre de motifs d'une des séquences dudit alignement, la valeur $A_{i,j}$ étant égale à une première valeur A_1 [par exemple "0"] lorsque le motif de position i de la séquence j est muté par rapport au motif de position i de la séquence de référence, et égale à une deuxième valeur A_2 [par exemple "1"] dans les autres cas,

25 - constituer deux matrices d'analyse B , C des mutations M où cette matrice est :

 - une matrice B de couples non mutés, c'est-à-dire de couples qui ne mutent jamais simultanément, de dimension $M \times M$, la valeur $B_{i,k} = B_{k,i}$ étant égale :

- 30 • à une première valeur B_1 [par exemple "0"] lorsque $A_{i,j} = A_{k,j} = 0$ quel que soit j allant de 0 à N ,
- à une deuxième valeur B_2 [par exemple "1"] dans les autres cas;

- une matrice C de couples mutés [c'est-à-dire de couples qui mutent soit au moins une fois simultanément, soit jamais] de dimension $M \times M$, la valeur $C_{i,k} = C_{k,i}$, étant égale :

- à une deuxième valeur C1 [par exemple "1"] lorsque $A_{i,j} = A_{k,j}$ quel que soit j allant de 0 à N,
- à une première valeur C2 [par exemple "0"] dans les autres cas;

- à déterminer, pour un ensemble E de positions, un coefficient R_E dont la valeur est R1 [par exemple "1"] lorsque toutes les valeurs $B_{i,k}$ sont égales à la deuxième valeur B2, quels que soient i et k appartenant à l'ensemble E desdites positions, où $i \neq j$.

- à déterminer, pour un ensemble F de positions, un coefficient R_F dont la valeur est R1 [par exemple "1"] lorsque toutes les valeurs $C_{i,k}$ sont égales à la deuxième valeur C2, quelles que soient i et k appartenant à l'ensemble F desdites positions, où $i \neq j$.

De préférence les séquences analysées par la méthode d'identification de l'invention est constitué par un sous-ensemble de séquences extrait d'une banque des séquences nucléotidiques ou polypeptidiques d'organismes pathogènes et tout préférentiellement par des séquences nucléotidiques ou polypeptidiques d'organismes pathogènes présentant un taux élevé de mutabilité.

Selon un mode de mise en oeuvre particulier le sous-ensemble de séquences comprend toutes les séquences polypeptidiques des différentes variantes connues de la protéase du virus de l'immunodéficience humaine.



Selon une autre mise en oeuvre particulière de l'invention le sous-ensemble de séquences comprend toutes les séquences polypeptidiques des différentes variantes connues de la transcriptase inverse du virus de l'immunodéficience humaine.

Selon une autre mise en oeuvre particulière de l'invention le sous-ensemble de séquences comprend toutes les séquences polypeptidiques des différentes variantes connues de l'intégrase du virus de l'immunodéficience humaine.

L'invention concerne l'identification de motifs appartenant à tout agent pathogène dont les séquences d'acides nucléiques et/ou polypeptidiques sont susceptibles de présenter des mutations.

A titre d'exemple de telles séquences et de manière non limitative, on peut citer les séquences de virus telles que le virus de l'hépatite C qui est un virus à ARN caractérisé par la grande variabilité de son génome, avec 3% de prévalence mondiale et 600 000 personnes infectées en France, les séquences du virus ébola qui provoque des fièvres hémorragiques et qui est associé à un fort taux de mortalité, les séquences du virus de la grippe pour lequel il est nécessaire de développer de nouveaux vaccins chaque année ou les séquences de tout autre virus émergeant à fort taux de mutabilité.

Ainsi, selon une mise en oeuvre particulière de l'invention le sous-ensemble de séquences extrait comprend toutes les séquences polypeptidiques des différentes variantes de la neuraminidase du virus de la grippe.

Selon une autre mise en oeuvre particulière de l'invention le sous-ensemble de séquences extrait comprend

toutes les séquences polypeptidiques des différentes variantes de l'hémagglutinine du virus de la grippe.

Aussi, parmi les séquences de bactéries susceptibles de présenter des mutations on peut également citer à titre d'exemple, la séquence C-terminal de la protéine HspA de la bactérie *Helicobacter Pylori* ou l'adhésine du type HA de la bactérie *Escherichia Coli*.

La méthode d'identification de motifs de l'invention n'est pas limitée au seul domaine des agents pathogènes. Des ensembles de séquences présentant des motifs n'ayant jamais muté simultanément, ou au contraire ayant muté simultanément au moins une fois sur au moins une des séquences de l'ensemble et n'ayant pas muté sur les autres séquences dudit ensemble, sont également présentes dans d'autres pathologies, comme par exemple, des pathologies dans le domaine de la cancérologie.

On admet, en effet qu'une grande partie des cancers est due à la présence d'éléments transposables ayant une grande homologie d'organisation avec les virus, et que le virus de l'hépatite B est la deuxième cause identifiée de mort par cancer après le tabac.

Aussi, parmi les gènes impliqués dans des cancers humains, susceptibles de présenter des motifs qui mutent et pour lesquels des ensembles de séquences ont parfois été constituées on peut citer à titre d'exemple : le gène APC , impliqué essentiellement dans le cancer du colon (Nucleic Acids Res 1998 Jan 1;26(1):269-70, APC gene: database of germline and somatic mutations in human tumors and cell lines. Laurent-Puig P, Beroud C, Soussi T.), le gène P53 (Nucleic Acids Res 1997 Jan 1;25(1):138 p53 and APC gene mutations: software and databases. Beroud C, Soussi T.), MEN-1 (A malignant gastrointestinal stromal tumour in a patient with multiple endocrine neoplasia type 1. Papillon E, Rolachon A, Calender A, Chabre O, Barnoud R,

Fournet J.), VHL (Mutations of the VHL gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. Gallou C, Joly D, Mejean A, Staroz F, Martin N, Tarlet G, Orfanelli MT, Bouvier R, Droz D, Chretien Y, Marechal JM, Richard S, Junien C, Beroud C.), WT1 (Clin Cancer Res 2000 Oct;6(10):3957-65. WT1 splicing alterations in Wilms' tumors. Baudry D, Hamelin M, Cabanis MO, Fournet JC, Tournade MF, Sarnacki S, Junien C, Jeanpierre C.)

L'invention a aussi pour objet l'utilisation de la méthode d'identification de motifs décrite ci-dessus pour la sélection de fragments de séquences constitués par et/ou comprenant des motifs n'ayant jamais muté simultanément pour la préparation de vaccins.

Les vaccins sont composés d'antigènes constitués par des molécules ou parties de molécules d'un organisme pathogène qui lorsqu'ils sont injectés dans l'organisme permettent de produire un plus grand nombre d'anticorps contre ledit organisme pathogène. Ces anticorps reconnaissent les molécules contre lesquelles ils sont dirigés et permettent ainsi au système immunitaire de détruire ledit organisme pathogène.

Or il s'écoule toujours un laps de temps non négligeable, parfois plusieurs années, entre le moment où l'on définit le vaccin et le moment où il arrive sur le marché. Par exemple en ce qui concerne le virus VIH, la faible fidélité de polymérisation de la reverse-transcriptase confère au virus une grande variabilité génomique qui augmente en fonction du temps. La population virale est ainsi très hétérogène. La destruction du virus

sauvage par le biais du vaccin conduit à la sélection des virus mutants contre lesquels le vaccin reste inefficace.

L'application de la méthode de l'invention à des sous-ensembles des séquences variantes des séquences protéiques de l'organisme pathogène permet de piéger ce dernier:

- Soit il mute, mais, dans ce cas, il n'est plus fonctionnel ;

- Soit, il ne mute pas, mais alors les anticorps produits à partir du vaccin permettront de le détruire.

Par exemple, en ce qui concerne le virus VIH, Les peptides faisant partie des protéines d'enveloppe du virus, identifiés parce qu'ils ne peuvent pas muter ensemble, probablement due à une pression génétique sous peine de perdre leur fonctionnalité, sont des candidats vaccins de choix.

En effet, la méthode d'identification de motifs peptidiques, permet de sélectionner de séquences contenant lesdits motifs, de manière contiguë ou non, afin d'élaborer un candidat vaccin. Ledit vaccin présente comme avantage, par rapport aux autres vaccins élaborées par des voies classiques, d'être décrit de façon exhaustive et de contenir de manière certaine les régions nécessaires à la stabilité dudit vaccin précisément par le choix des séquences ne pouvant pas muter ensemble simultanément, entraînant ainsi la destruction de l'organisme pathogène.

L'identification des motifs n'ayant jamais muté simultanément est plus complexe pour deux raisons principales :

- Le nombre d'acides aminés ne mutant jamais est à peu près dix fois plus grand,
- La combinaison d'acides aminés à tester n'étant pas déterminée à l'avance, toutes les combinaisons doivent être envisagées.

L'invention concerne également l'utilisation des fragments de séquences constitués par et/ou comprenant des motifs nucléotidiques et/ou peptidiques des séquences analysées, n'ayant jamais muté simultanément pour la préparation de vaccins.

L'invention a également pour objet l'utilisation d'une telle méthode d'identification de motifs ou de combinaison de motifs n'ayant jamais muté simultanément pour la conception d'outils d'aide au diagnostic.

En effet, la méthode de l'invention permet également de construire une base de connaissances qui constitue un outil d'aide à la décision, par exemple lors de la détermination par le médecin de l'administration des traitements anti-viraux à un patient donné.

Selon un autre mode de mise en œuvre de l'invention, la méthode d'identification de motifs n'ayant jamais muté simultanément comprend une étape supplémentaire consistant à comparer des données reliant les résistances médicamenteuses connues aux mutations observées, par exemple dans les cas du VIH aux données divulguées par J.

Hammond et al. dans " Mutations in Retroviral Genes Associated with Drug Resistance ". (The Human Retroviruses and AIDS Compendium. 1999)

5 La relation drogue-acide aminé muté, ainsi mise en évidence, est très utile pour optimiser le traitement. Par exemple, en ce qui concerne le virus VIH, la comparaison des motifs peptidiques s'effectue sur trois sous-ensembles d'une base de données protéiques, celui de la reverse transcriptase, celui de la protéase et celui de
10 l'intégrase (<http://hiv-web.lanl.gov/>).

 La comparaison des séquences appartenant audits sous-ensembles comprenant de 300 à 8000 séquences, ou de
15 fragments desdites séquences, de chacune de ces trois protéines, permet en appliquant la méthode de l'invention, d'identifier des combinaisons d'acides aminés qui n'ont jamais muté simultanément.

 Ainsi, la méthode de l'invention permet alors
20 d'identifier les mutations induites sous la pression de sélection.

 Alors la méthode de l'invention, comprenant la comparaison avec lesdites résistances médicamenteuse permet de choisir une combinaison de drogues de manière à ce que
25 les mutations d'acides aminés susceptibles d'être provoquées par chacun des antiviraux, susceptibles de conférer la résistance aux différents médicaments impliqués dans cette combinaison (moins d'une dizaine), ne se produisent pas simultanément.

30

 L'identification de tels motifs permet la sélection d'une combinaison médicamenteuse qui défavorise

l'apparition de plus d'une mutation à la fois fermant ainsi la porte aux plurirésistances.

Le praticien pourra ensuite utiliser les informations obtenues en appliquant cette méthode par exemple aux séquences virales isolées, ou déduites du génome viral isolé, d'un patient donné pour s'assurer que la multithérapie envisagée est en effet la plus efficace possible.

L'identification d'une première mutation excluant les deux autres, une trithérapie ainsi choisie permet aux deux composés antirétroviraux restant de demeurer efficaces.

La méthode d'identification de régions peptidiques n'ayant pas muté simultanément selon l'invention apporte également une aide précieuse lors de l'apparition de résistances chez des malades déjà traités.

La méthode selon l'invention peut par exemple s'appliquer à des sous-ensembles de séquences polypeptidiques parmi lesquelles est incluse celle ou celles déduites à partir du séquençage du génome viral isolé du patient.

Ainsi, si ce génotypage met en évidence une mutation responsable de la résistance, la méthode d'identification de motifs peptidiques n'ayant pas muté permet de mettre en oeuvre une multithérapie conçue de manière à maintenir la pression de sélection sur la mutation.

La molécule ainsi sélectionnée sera accompagnée de deux ou trois autres antirétroviraux qui ciblent des domaines de la protéine ne pouvant pas muter en même temps que la zone ayant déjà muté.

Une telle méthode est utile pour la mise en oeuvre de nouvelles combinaisons antirétrovirales empêchant au maximum l'échappement thérapeutique.

5 Aussi, par exemple, l'identification de motifs, à l'intérieur d'un même gène, ayant muté au moins une fois simultanément sur au moins une variante et n'ayant jamais muté sur les autres variantes, permet d'identifier des régions dudit gène susceptibles de présenter une
10 interaction physique ou fonctionnelle. En revanche, l'identification des motifs n'ayant jamais muté simultanément permet d'identifier des régions dudit gène dont la présence mutuelle est essentielle et indispensable à sa fonction.

15 L'invention a également pour objet l'identification, sur un ensemble de gènes ou sur un ensemble de séquences non-codantes, de motifs n'ayant jamais muté simultanément. L'identification de tels motifs
20 permet de sélectionner des régions géniques susceptibles de présenter des interactions physiques ou fonctionnelles sur l'ensemble du génome.

25 Un autre objet de l'invention concerne l'utilisation d'une telle méthode d'identification de motifs ou de combinaisons de motifs pour la sélection de fragments de séquences constitués et/ou comprenant des motifs n'ayant jamais muté simultanément pour la
30 préparation de cibles thérapeutiques.

Encore un autre objet de l'invention se rapporte à l'utilisation de fragments de séquences constitués par et/ou comprenant des motifs soit ayant muté

simultanément au moins une fois sur au moins une séquence de l'ensemble et n'ayant jamais muté sur les autres séquences de l'ensemble pour la préparation de cibles thérapeutiques.

5

L'invention a également pour objet l'utilisation des motifs ou combinaisons de motifs ainsi identifiés pour préparer des cibles thérapeutiques utiles pour le criblage de nouveaux composés thérapeutiques destinés à la prévention et/ou traitement des pathologies humaines, animales ou végétales.

10

Ainsi, la préparation, après avoir identifié des motifs n'ayant jamais muté simultanément, de fragments de séquence les contenant, permet la préparation d'une cible thérapeutique contre laquelle seront testés des composés thérapeutiques dirigés contre ledit organisme pathogène et notamment des composés thérapeutiques contre lesquels l'organisme pathogène sauvage ne pourra pas développer de mutations de résistance.

15

20

La sélection de fragments constitués et/ou comportant de motifs n'ayant jamais muté simultanément est aussi utile pour la préparation d'outils de diagnostic où il n'est pas toujours facile de détecter rapidement tel ou tel type ou sous-type d'organisme pathogène, car l'identification de motifs peptidiques selon l'invention permet la préparation des fragments de peptides comprenant les motifs les plus représentatifs d'un sous-type d'organisme pathogène. Ces fragments sont ensuite utilisés dans des tests de détection, comme des tests immunoenzymatiques, par exemple.

25

30

Cette application de la méthode de l'invention consiste à identifier un ensemble de motifs indispensable à la fonction d'une protéine d'un organisme humain, animal ou végétal ou d'un organisme pathogène. Ces motifs peuvent constituer, par exemple, un sous-ensemble d'acides aminés connus pour jouer un rôle important dans la fonction de la protéine ciblée.

Avantageusement, les motifs ainsi identifiés sont des motifs contigus de la séquence génique et représentent une séquence linéaire dudit gène.

Avantageusement, les motifs identifiés sont des motifs non contigus sur la séquence linéaire du gène. Ils peuvent alors être utiles pour compléter des études d'analyse tridimensionnelle afin de confirmer une éventuelle proximité spatiale non linéaire desdits motifs. La méthode de l'invention peut comporter alors une nouvelle étape supplémentaire (g), après l'étape (e) d'identification des motifs, consistant à comparer lesdits motifs avec les données de structures tridimensionnelles de ces protéines tels que des acides aminés impliqués dans le site catalytique et/ou dans les sites liés par des inhibiteurs non-compétitifs.

Cette dernière comparaison fournit une liste d'acides aminés impliqués dans la fonction protéique et ne mutant jamais ensemble.

La méthode d'identification de régions peptidiques selon l'invention définit les peptides les plus représentatifs d'un sous-type. Une fois identifiés, ces peptides sont utilisés dans tout test de détection connu de

l'homme du métier, tels que des tests immunoenzymatiques, du type ELISA, par exemple.

La recherche de peptides représentant un sous-type d'un type particulier s'effectue comme indiqué ci-dessus. Il s'agit de trouver des antigènes peptidiques capables d'être reconnus par un sérum particulier contenant ou non les anticorps d'un sous-type particulier. La méthode selon l'invention peut s'appliquer à n'importe quelle banque de séquences, les résultats sont comparés par sous-types et la combinaison peptidique théorique la plus représentative d'un type pathogène particulier est ainsi identifiée.

Les peptides ainsi identifiés sont synthétisés et testés immunologiquement contre une collection de sérums.

L'invention présente tout son intérêt lorsqu'elle est utilisée pour identifier soit des motifs ayant muté au moins une fois ensemble, soit n'ayant jamais muté à partir d'un grand nombre de séquences comportant un grand nombre de motifs afin de sélectionner des séquences de motifs utiles pour les différentes applications envisagées ci-dessus.

Afin d'illustrer la méthode d'identification de motifs de l'invention, l'exemple ci-après montre les différentes matrices constituées lors d'une comparaison de motifs effectuée sur un sous-ensemble de huit séquences, à l'aide de la séquence de référence S V R L G H K D E V.

POSITIONS	0 1 2 3 4 5 6 7 8 9
Séquence de référence (consensus)	S V R L G H K D E V

Sous-ensemble de séquences	Alignement
Seq 1	S R R L G H K D E V
Seq 2	S V R L G H K L E V
Seq 3	S R D L G H K D E V
Seq 4	S V R L G H L D V V
Seq 5	S V D L G H K T E V
Seq 6	S K R L G H K D E V
Seq 7	S V R L G H G D G V
Seq 8	S V R L G H K S E V

5 1 - MATRICE DE MUTATION A.

Valeurs attribuées :

10 A1=0, si motif muté par rapport à la séquence de référence
A2=1, si autre cas (motif non muté par rapport à la séquence de référence).

POSITION	0 1 2 3 4 5 6 7 8 9
Seq 1	1 0 1 1 1 1 1 1 1 1
Seq 2	1 1 1 1 1 1 1 0 1 1
Seq 3	1 0 0 1 1 1 1 1 1 1
Seq 4	1 1 1 1 1 1 0 1 0 1
Seq 5	1 1 0 1 1 1 1 0 1 1
Seq 6	1 0 1 1 1 1 1 1 1 1
Seq 7	1 1 1 1 1 1 0 1 0 1
Seq 8	1 1 1 1 1 1 1 0 1 1

L'interrogation de la matrice mutée C permet ainsi d'identifier les motifs en positions 6 et 8 comme des motifs ayant muté au moins une fois ensemble.

REVENDICATIONS

1) Méthode d'identification d'un motif ou d'une
5 combinaison de motifs présentant un état booléen de
mutations prédéterminées dans un ensemble de séquences,
caractérisée en ce qu'elle comprend au moins les étapes
suivantes :

a) l'alignement de l'ensemble de séquences de
10 motifs ordonnés représentées par leur code unicaractère,

b) la comparaison d'une séquence de référence à
l'ensemble de séquences alignées à l'étape (a),

c) l'identification des motifs n'ayant jamais
muté simultanément ou au contraire des motifs ayant muté
15 simultanément au moins une fois sur au moins une des
séquences de l'ensemble et n'ayant pas muté sur les autres
séquences dudit ensemble.

2) Méthode d'identification , selon la
20 revendication 1 caractérisée en ce que le motif ou la
combinaison de motifs est un nucléotide ou une combinaison
de nucléotides et en ce que le sous-ensemble de séquences
est choisi parmi les séquences d'une banque de donnés
d'acides nucléiques.

25 3) Méthode d'identification selon la
revendication 1 caractérisée en ce que le motif ou la
combinaison de motifs est un acide aminé ou une combinaison
d'acides aminés et en ce que le sous-ensemble de séquences
30 est choisi parmi les séquences d'une banque de donnés de
polypeptides et/ou de protéines.

4) Procédé d'identification selon l'une des revendications 1 à 3, caractérisé en ce que la séquence de référence utilisée pour la comparaison de l'étape (b) est une séquence sauvage.

5

5) Procédé d'identification selon l'une des revendications 1 à 3, caractérisé en ce que la séquence de référence utilisée pour la comparaison de l'étape (b) est une séquence comportant en position i le motif présent en position i dans un nombre prédéterminé des séquences de l'étape (a), par exemple dans plus de 30% desdites séquences et plus préférentiellement dans plus de 75% desdites séquences.

10

6) Méthode d'identification selon l'une des revendications 1 à 5, caractérisée en ce que l'étape (b) de comparaison de séquences consiste à :

15

- constituer une première matrice numérique A de dimensions $N \times M$ où N désigne le nombre de séquences et M désigne le nombre de motifs d'une des séquences dudit alignement, la valeur $A_{i,j}$ étant égale à une première valeur $A1$ lorsque le motif de position i de la séquence j est muté par rapport au motif de position i de la séquence de référence, et égale à une deuxième valeur $A2$ dans les autres cas,

20

- constituer deux matrices d'analyse B , C des mutations où cette matrice est :

25

- une matrice B de couples non mutés, c'est-à-dire de couples qui ne mutent jamais simultanément, de dimension $M \times M$, la valeur $B_{i,k} = B_{k,i}$ étant égale :

30

- à une première valeur $B1$ lorsque $A_{i,j} = A_{k,j} = A1$ quel que soit j allant de 0 à N ,
- à une deuxième valeur $B2$ dans les autres cas;

- une matrice C de couples mutés de dimension $M \times M$, la valeur $C_{k,i} = C_{i,k}$ étant égale :

- à une deuxième valeur C1 lorsque $A_{i,j} = A_{k,j}$ quel que soit j allant de 0 à N,
- à une première valeur C2 dans les autres cas;

- déterminer, pour un ensemble E de positions, un coefficient R_E dont la valeur est R1 lorsque toutes les valeurs $B_{i,k}$ sont égales à la deuxième valeur B2, quels que soient i et k appartenant à l'ensemble E desdites positions, où $i \neq j$.

- déterminer, pour un ensemble F de positions, un coefficient R_F dont la valeur est R1 lorsque toutes les valeurs $C_{i,k}$ sont égales à la deuxième valeur C2, quels que soient i et k appartenant à l'ensemble F desdites positions, où $i \neq j$.

7) Méthode d'identification selon l'une quelconque des revendications 1 à 6, caractérisée en ce que à l'étape (c), les positions des ensembles E et/ou F sont désignées par l'utilisateur

8) Méthode d'identification selon l'une quelconque des revendications 1 à 6 caractérisée en ce que l'étape (c) comporte une étape de test consistant à générer la totalité des combinaisons de positions possibles et à déterminer pour chacune desdites combinaisons la valeur des coefficients R_E ou R_F , et à retenir la combinaison correspondant au plus grand ensemble de positions dont le coefficient R_E ou R_F correspond à ladite deuxième valeur.

9) Méthode d'identification selon l'une quelconque des revendications 1 à 8 caractérisée en ce que l'ensemble de séquences analysées est constitué par des

séquences de motifs d'organismes pathogènes et de préférence d'organismes pathogènes présentant un taux élevé de mutabilité.

5 10) Méthode d'identification selon l'une
quelconque des revendications 1 à 8 caractérisée en ce que
l'ensemble de séquences analysées est constitué par des
séquences de motifs de gènes, impliqués dans des
pathologies humaines, animales ou végétales et de
10 préférence présentant un taux élevé de mutabilité.

11) Utilisation de la méthode d'identification
de motifs selon l'une quelconque des revendications 1 à 10
pour la sélection de fragments de séquence constitués par
15 et/ou comprenant des motifs n'ayant jamais muté
simultanément pour la préparation de vaccins.

12) Utilisation de la méthode d'identification
de motifs selon les revendications 1 à 10 pour la sélection
20 de fragments de séquence constitués par et/ou comprenant
des motifs n'ayant jamais muté simultanément pour la
préparation de cibles thérapeutiques

13) Utilisation de la méthode d'identification
25 selon les revendications 1 à 10 pour la sélection pour la
sélection de fragments de séquence constitués par et/ou
comprenant des motifs n'ayant jamais muté simultanément
pour la préparation de tests de diagnostic.

30 14) Utilisation de la méthode d'identification
selon les revendications 1 à 10 pour la sélection de
fragments de séquence constitués par et/ou comprenant des

motifs ayant toujours muté simultanément pour la préparation de tests de diagnostic.

5 15) Utilisation des fragments de séquence constitués par et/ou comprenant des motifs nucléotidiques et/ou peptidiques, n'ayant jamais muté simultanément pour la conception de vaccins.

10 16) Utilisation de fragments de séquence constitués par et/ou comprenant des motifs n'ayant jamais muté simultanément pour la préparation de cibles thérapeutiques utiles pour le criblage de nouveaux composés thérapeutiques destinés à la prévention et/ou traitement des pathologies humaines animales ou végétales.

15 17) Utilisation de fragments de séquence constitués par et/ou comprenant des motifs ayant toujours muté simultanément pour la préparation de cibles thérapeutiques utiles pour le criblage de nouveaux composés thérapeutiques destinés à la prévention et/ou traitement des pathologies humaines, animales ou végétales.

25 18) Utilisation de fragments de séquence constitués par et/ou comprenant des motifs n'ayant jamais muté simultanément pour la conception d'outils d'aide au diagnostic.

30 19) Utilisation selon l'une quelconque des revendications 15 ou 16, caractérisée en ce que les fragments de séquence sélectionnés comprennent des motifs n'ayant jamais muté simultanément contigus.

20) Utilisation selon l'une quelconque des revendications 14 ou 17, caractérisée en ce que les fragments de séquence sélectionnés comprennent des motifs ayant muté simultanément contigus.

5

21) Utilisation selon l'une quelconque des revendications 15 ou 16, caractérisée en ce que les fragments de séquence sélectionnés comprennent des motifs n'ayant jamais muté simultanément non-contigus.

10

22) Utilisation selon l'une quelconque des revendications 14 ou 17, caractérisée en ce que les fragments de séquence sélectionnés comprennent des motifs ayant muté simultanément non-contigus.

15

23) Utilisation de fragments de séquences constitués par et/ou comprenant des motifs ayant toujours muté simultanément pour la conception d'outils d'aide au diagnostic.

20

24) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 ou 3 à 10 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences polypeptidiques des différentes variantes de la protéase du virus de l'immunodéficience humaine.

25

25) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 ou 3 à 9 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences polypeptidiques des différentes variantes de la transcriptase inverse du virus de l'immunodéficience humaine.

30

26) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 ou 3 à 9 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences polypeptidiques des différentes variantes de l'intégrase du virus de l'immunodéficience humaine.

27) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 à 10 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences de motifs des différentes variantes du gène ou de la protéine de la neuraminidase du virus de la grippe.

28) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 à 10 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences de motifs des différentes variantes du gène ou de la protéine de l'hémagglutinine du virus de la grippe.

29) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 à 10 caractérisée en ce que l'ensemble de séquences de l'étape (a) comprend toutes les séquences de motifs des différentes variantes d'un gène et/ou d'une protéine du virus de l'hépatite C.

30) Méthode d'identification de motifs, selon l'une quelconque des revendications 1 à 10 caractérisée en ce que l'ensemble de séquences de motifs de l'étape (a) comprend toutes les séquences des différentes variantes des

séquences du gène ou de la protéine HspA de la bactérie *Helicobacter Piloni*.

31) Méthode d'identification de motifs, selon
5 l'une quelconque des revendications 1 à 10, caractérisée en
ce que le sous-ensemble de séquences de motifs sélectionnée
à l'étape (a) comprend toutes les séquences des différentes
variantes du gène ou de la protéine de l'adhésine du type
HA de la bactérie *Escherichia Coli*

10 32) Méthode d'identification de motifs, selon
l'une quelconque des revendications 1 à 10, caractérisée en
ce qu'elle comprend, après l'étape (c), une étape
supplémentaire (d) de comparaison de motifs identifiés lors
15 de ladite étape (c) avec les résistances médicamenteuses
connues aux mutations observées.

20 33) Méthode d'identification de motifs, selon
l'une quelconque des revendications 1 à 10, caractérisé en
ce qu'elle comprend, après l'étape (c), une étape
supplémentaire (e) de comparaison de motifs identifiés lors
de ladite étape (c) avec des motifs des séquences impliqués
dans un site catalytique et/ou dans des sites liés par des
inhibiteurs non-compétitifs.

25

DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

DÉSIGNATION D'INVENTEUR(S) Page N° 1. / 1.

(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

08 113 W / 260899

Vos références pour ce dossier (facultatif)		13052FR	
N° D'ENREGISTREMENT NATIONAL		01/07808	
TITRE DE L'INVENTION (200 caractères ou espaces maximum) Méthode d'identification de motifs et/ou des combinaisons de motifs présentant un état booleen de mutation prédéterminée dans un ensemble de séquences et ses applications.			
LE(S) DEMANDEUR(S) : Centre National de la Recherche Scientifique -CNRS- 3, rue Michel Ange 75794 PARIS Cedex 16 Institut Pasteur 25-28, rue du Docteur ROUX 75724 PARIS Cedex 15			
DESIGNE(NT) EN TANT QU'INVENTEUR(S) : (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		VANET	
Prénoms		Anne	
Adresse	Rue	52, rue de Crimée	
	Code postal et ville	75019	PARIS
Société d'appartenance (facultatif)			
Nom		MULLER-TRUTWIN	
Prénoms		Michaela	
Adresse	Rue	110, rue Vieille du Temple	
	Code postal et ville	75003	PARIS
Société d'appartenance (facultatif)			
Nom		VALERE	
Prénoms		Thomas	
Adresse	Rue	36, rue des Roses	
	Code postal et ville	75018	PARIS
Société d'appartenance (facultatif)			
DATE ET SIGNATURE(S) DU (DES) DEMANDEUR(S) OU DU MANDATAIRE (Nom et qualité du signataire) Le 12/02/02 BREESE Pierre 921038			

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.